

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2001, 09:00:24 ; Search time 9.26 Seconds  
(without alignments)  
31.387 Million cell updates/sec

Title: US-09-307-223-1  
Perfect score: 57  
Sequence: 1 CRRETAMAC 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues  
Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41	71.9	815	1	GYRB_MYXXA
2	37	64.9	307	1	YD57_MYCTU
3	36	63.2	225	1	TRMA_NEIGO
4	35	61.4	283	1	LECH_MOUSE
5	35	61.4	283	1	LECH_MOUSE
6	35	61.4	347	1	GALE_RAT
7	35	61.4	417	1	PVR_HUMAN
8	35	61.4	501	1	VNS1_AHSV9
9	34	59.6	501	1	PHR1_SIGNAL
10	33	57.9	251	1	PIG4_STNEL
11	33	57.9	262	1	VME1_CVPFS
12	33	57.9	536	1	PPCK_RHIME
13	33	57.9	838	1	HIS2_CANAL
14	33	57.9	939	1	XPC_HUMAN
15	32	56.1	255	1	HA25_HUMAN
16	32	56.1	301	1	LECI_RAT
17	32	56.1	350	1	PRIM_METUA
18	32	56.1	374	1	YDVA_SCHPO
19	32	56.1	537	1	IL2B_RAT
20	32	56.1	539	1	IL2B_MOUSE
21	32	56.1	665	1	PD12_HUMAN
22	32	56.1	1533	1	PDK2_DROME
23	31	54.4	136	1	HIS3_METVA
24	31	54.4	166	1	YK22_GVXN
25	31	54.4	179	1	YKHA_BACSU
26	31	54.4	201	1	GSHC_HUMAN
27	31	54.4	257	1	KDRX_SERMA
28	31	54.4	277	1	J11_HCMVA
29	31	54.4	290	1	LECH_HUMAN
30	31	54.4	306	1	MMGL_RAT
31	31	54.4	327	1	YC62_PORPU
32	31	54.4	382	1	Y022_NPVAC
33	31	54.4	382	1	Y022_NPVOP

34	31	54.4	428	1	MLH1_ALCEU
35	31	54.4	435	1	SL52_BRAOA
36	31	54.4	533	1	IMAS_MOUSE
37	31	54.4	534	1	AMT1_CAEL
38	31	54.4	536	1	IMAS_HUMAN
39	31	54.4	537	1	PPCK_RHISN
40	31	54.4	538	1	IMAI_HUMAN
41	31	54.4	538	1	IMAI_MOUSE
42	31	54.4	548	1	VNS1_AHSV4
43	31	54.4	548	1	VNS1_AHSV6
44	31	54.4	551	1	YK27_YEAST
45	31	54.4	666	1	ENV_MLWHO

## ALIGNMENTS

RESULT 1  
ID GYRB\_MYXXA STANDARD: PRT: 815 AA.  
AC 03367;  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE DNA GYRASE SUBUNIT B (EC 5.99.1.3).  
GN GYRB.  
OS Myxococcus xanthus.  
OC Bacteria; Proteobacteria; delta subdivision; Myxobacteria;  
OC Myxococcales; Cytophacterales; Myxococcaceae; Myxococcus.  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN-ER-15;  
RA Paltan Y., Boulton N., Ron E., Rosenberg E., Orr E.,  
RL Submitted (Jun-1997) to the EMBL/GenBank/DDJ closed databases.  
CC -1- FUNCTION: DNA GYRASE NEGATIVELY SUPERCOILS CLOSED CIRCULAR DOUBLE-  
CC INTERCONVERSION OF OTHER TOPOLOGICAL ISOMERS OF DOUBLE-STRANDED  
CC DNA RINGS, INCLUDING CATENANES AND KNOTTED RINGS.  
CC -1- CATALYTIC ACTIVITY: ATP-DEPENDENT BREAKAGE, PASSAGE AND REJOINING  
CC OF DOUBLE-STRANDED DNA.  
CC -1- SUBUNIT: MADE UP OF TWO CHAINS. THE A CHAIN IS RESPONSIBLE FOR DNA  
CC BREAKAGE AND REJOINING; THE B CHAIN CATALYZES ATP HYDROLYSIS. THE  
CC ENZYME FORMS AN A2B2 TETRAMER.  
CC -1- SIMILARITY: BELONGS TO TYPE II TOPOISOMERASE FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC  
CC EMBL: AJ000543; CA04176.1; -  
CC INTERPRO: IPR000565; -  
CC INTERPRO: IPR001241; -  
CC INTERPRO: IPR002288; -  
CC PFM: PF000986; DNA\_GYRASE\_B; 1.  
CC PFM: PF00204; DNA\_TOPOISOM; 1.  
CC PRINTS: PR01159; DNAGYRASE.  
CC DR PROSITE: PS00177; TOPOISOMERASE-II; 1.  
CC KW Topoisomerase; Isomerase; ATP-binding.  
CC SEQUENCE 815 AA; 89636 MW; 3862685FBB805B32 CRC64;

Query Match 71.9%; Score 41; DB 1; Length 815;  
Best Local Similarity 55.6%; Pred. No. 4;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 CRRETAMAC 9  
DB 693 CRRSTWSC 701

```

RESULT 2
YD57_MYCTU STANDARD; PRT; 307 AA.
ID YD57_MYCTU
AC 011027;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HYPOETHERICAL 33.9 KDA PROTEIN RV1357C.
GN RV1357C OR MTCY02B10.21C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-98295987; PubMed-9634230;
RA Cole S.T., Broch R., Parkhill J., Gaiter T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornby T., Jagsen K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrett B.G.;
RT Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.
RL Nature 393:537-544(1998).
CC -1- SIMILARITY: BELONGS TO THE YEGE/YHDA/YHJK/YJOC FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: 275555; CA93962.1;
CC DR TUBERCULIST; RV1357C;
CC DR INTERPRO: IPR001633;
CC DR PFAM: PF00563; DUF2; 1.
CC KM Hypothetical protein; Transmembrane.
CC FT TRANSMEM 158 178
CC FT TRANSMEM 203 223
CC FT TRANSMEM 203 223
CC FT TRANSMEM 203 223
CC SEQUENCE 307 AA; 33982 MW; 3DC5976AB3BAC48 CRC64;

```

```

Query Match 64.9%; Score 37; DB 1; Length 307;
Best Local Similarity 66.7%; Pred. No. 8.2;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 CRRETAMAC 9
DB 5 CORATAFAC 13

```

```

RESULT 3
TRMA_NEIGO STANDARD; PRT; 225 AA.
ID TRMA_NEIGO
AC P55134;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE PROBABLE TRNA (URACIL-5)-METHYLTRANSFERASE (EC 2.1.1.35) (TRNA/M-5-
DE USA)-METHYLTRANSFERASE (RUMT) (FRAGMENT).
GN TRMA.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-93181472; PubMed-8441767;
RA Sarandopoulos S., Davies J.K.;

```

```

RT "Lack of substantial sequence homology between the cryptic plasmid
RT and chromosome of Neisseria gonorrhoeae."
RL Plasmid 29:41-49(1993).
CC -1- FUNCTION: CATALYZES THE FORMATION OF 5-METHYL-URIDINE AT POSITION
CC 54 (M-5-U54) IN ALL TRNA (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: S-ADENOSYL-L-METHIONINE + TRNA - S-ADENOSYL-L-
CC L-HOMOCYSTEINE + TRNA CONTAINING THYMINE.
CC -1- SIMILARITY: BELONGS TO THE RNA METHYLTRANSFERASE TRNA FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M90807; AAA63404.1;
CC DR INTERPRO: IPR001566;
CC DR PROSITE: PS01230; TRMA_1; PARTIAL.
CC DR PROSITE: PS01231; TRMA_2; PARTIAL.
CC KM Transferase; Methyltransferase; TRNA processing.
CC FT NON_TER 1 1
CC FT NON_TER 1 1
CC SEQUENCE 225 AA; 25680 MW; 9AB84BA4C8B7215B CRC64;

```

```

Query Match 63.2%; Score 36; DB 1; Length 225;
Best Local Similarity 55.6%; Pred. No. 9.1;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 1 CRRETAMAC 9
DB 120 CRKLEMAC 128

```

```

RESULT 4
LECH_MOUSE STANDARD; PRT; 283 AA.
ID LECH_MOUSE
AC P34927; O64363;
DT 01-FEB-1994 (Rel. 28, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ASIALOGLYCOPROTEIN RECEPTOR 1 (HEPATIC LECTIN 1) (MHL-1) (ASGP-R)
DE (ASGPR).
GN ASGPR OR ASGR-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-93176818; PubMed-8439566;
RX Takekawa R., Shinzawa K., Watanabe Y., Akaike T.;
RT "Determination of mouse major asialoglycoprotein receptor cDNA
RT sequence."
RT Blochim. Biophys. Acta 1172:220-222(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN-BALB/C; TISSUE-LIVER;
RX MEDLINE-95047431; PubMed-7958950;
RA Monroe R.S., Huber B.E.;
RT "The major form of the murine asialoglycoprotein receptor: cDNA
RT sequence and expression in liver, testis and epididymis."
RL Gene 146:237-244(1994).
CC -1- FUNCTION: MEDIATES THE ENDOCYTOSIS OF PLASMA GLYCOPROTEINS TO
CC WHICH THE TERMINAL SIALIC ACID RESIDUE ON THEIR COMPLEX
CC CARBOHYDRATE MOIETIES HAS BEEN REMOVED. THE RECEPTOR RECOGNIZES
CC TERMINAL GALACTOSE AND N-ACETYLGLUCOSAMINE UNITS. AFTER LIGAND
CC BINDING TO THE RECEPTOR, THE RESULTING COMPLEX IS INTERNALIZED AND
CC TRANSPORTED TO A SORTING ORGANELLE, WHERE RECEPTOR AND LIGAND ARE
CC DISASSOCIATED. THE RECEPTOR THEN RETURNS TO THE CELL MEMBRANE
CC SURFACE.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.

```

```

CC -1- TISSUE SPECIFICITY: EXPRESSED EXCLUSIVELY IN HEPATIC PARENCHYMAL
CC CELLS.
CC -1- MISCELLANEOUS: CALCIUM IS REQUIRED FOR LIGAND BINDING.
CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-sdb.ch).
CC -----
DR EMBL: D13517; BAA02734.1;
DR EMBL: U09362; AAB60440.1;
DR EMBL: U08372; AAB60440.1;
DR PIR: S29855; S29855.
DR HSSP: P20693; 1HLJ.
DR MGD: MGI:88081; ASGR1.
DR INTERPRO: IPR001304;
DR PFAM: PF00059; lectin_c_1.
DR PROSITE: PS00615; C-TYPE_LECTIN_1; 1.
DR PROSITE: PS50041; C-TYPE_LECTIN_2; 1.
KW Lectin; Glycoprotein; Receptor; Endocytosis; Transmembrane;
KW Calcium; Signal-anchor; Phosphorylation.
FT INIT MET 0 0
FT DOMAIN 1 38
FT TRANSMEM 39 59
FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT EXTRACELLULAR (POTENTIAL).
FT DOMAIN 151 277
FT SITE 4 7
FT C-TYPE LECTIN (LONG FORM).
FT ENDOCYTOSIS SIGNAL (POTENTIAL).
FT DISULFID 152 163
FT BY SIMILARITY.
FT DISULFID 180 275
FT BY SIMILARITY.
FT DISULFID 253 267
FT BY SIMILARITY.
FT CARBOHYD 74 74
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 77 77
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 145 145
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 150 150
FT I -> T (IN REF. 1).
SQ SEQUENCE 283 AA; 32472 MW; 982A5D305AAED08F CRC64;

Query Match 61.4%; Score 35; DB 1; Length 283;
Best Local Similarity 55.6%; Pred. No. 17;
Matches 5; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 CRRETAMAC 9
DB 267 CRPYRWVC 275

RESULT 5
LECH_RAT STANDARD; PRT; 283 AA.
ID LECH_RAT
AC P02706;
DT 21-JUL-1986 (Rel. 01, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 30-MAY-2000 (Rel. 35, Last annotation update)
DE ASIALOGLYCOPROTEIN RECEPTOR 1 (HEPATIC LECTIN 1) (RHL-1) (ASGP-R)
DE (ASGP-R).
GN ASGR1 OR ASGR-1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-86008335; PubMed-2995379;
RX Leung J.O., Holland E.C., Drickamer K.;
RT "Characterization of the gene encoding the major rat liver
RT asialoglycoprotein receptor.";
RL J. Biol. Chem. 260:12523-12527(1985).
RN [2]
RP SEQUENCE FROM N.A.

```

```

RX MEDLINE-85063786; PubMed-6095287;
RX Holland E.C., Leung J.O., Drickamer K.;
RT "Rat liver asialoglycoprotein receptor lacks a cleavable NH2-terminal
RT signal sequence.";
RL Proc. Natl. Acad. Sci. U.S.A. 81:7338-7342(1984).
RN [3]
RP SEQUENCE OF 11-283 FROM N.A.
RX MEDLINE-87026895; PubMed-2945599;
RX Watts C.;
RT "Isolation and expression of cDNA clones for a rat liver
RT asialoglycoprotein receptor.";
RL Biosci. Rep. 6:527-534(1986).
CC -1- FUNCTION: MEDIATES THE ENDOCYTOSIS OF PLASMA GLYCOPROTEINS TO
CC WHICH THE TERMINAL SIALIC ACID RESIDE ON THEIR COMPLEX
CC CARBOHYDRATE MOITIES HAS BEEN REMOVED. THE RECEPTOR RECOGNIZES
CC TERMINAL GALACTOSE AND N-ACETYLGLUCOSAMINE UNITS. AFTER LIGAND
CC BINDING TO THE RECEPTOR, THE RESULTING COMPLEX IS INTERNALIZED AND
CC TRANSPORTED TO A SORTING ORGANELLE, WHERE RECEPTOR AND LIGAND ARE
CC DISASSOCIATED. THE RECEPTOR THEN RETURNS TO THE CELL MEMBRANE
CC SURFACE.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: EXPRESSED EXCLUSIVELY IN HEPATIC PARENCHYMAL
CC CELLS.
CC -1- MISCELLANEOUS: CALCIUM IS REQUIRED FOR LIGAND BINDING.
CC -1- MISCELLANEOUS: TWO TYPES OF RAT HEPATIC LECTIN HAVE BEEN
CC IDENTIFIED, RHL-1 AND RHL-2/3. HAVING A RELATIVE ABUNDANCE OF 4:1.
CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-sdb.ch).
CC -----
DR EMBL: K02817; AAA42037.1;
DR EMBL: M21770; AAA40764.1;
DR PIR: A03166; LNRL.
DR HSSP: P20693; 1HLJ.
DR INTERPRO: IPR001304;
DR PFAM: PF00059; lectin_c_1.
DR PROSITE: PS00615; C-TYPE_LECTIN_1; 1.
DR PROSITE: PS50041; C-TYPE_LECTIN_2; 1.
KW Lectin; Glycoprotein; Receptor; Endocytosis; Transmembrane;
KW Calcium; Signal-anchor; Phosphorylation.
FT INIT MET 0 0
FT DOMAIN 1 38
FT TRANSMEM 39 59
FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT EXTRACELLULAR (POTENTIAL).
FT DOMAIN 151 277
FT SITE 4 7
FT C-TYPE LECTIN (LONG FORM).
FT ENDOCYTOSIS SIGNAL (POTENTIAL).
FT DISULFID 152 163
FT BY SIMILARITY.
FT DISULFID 180 275
FT BY SIMILARITY.
FT DISULFID 253 267
FT BY SIMILARITY.
FT CARBOHYD 74 74
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 77 77
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 145 145
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 60 60
FT O -> R (IN REF. 2).
SQ SEQUENCE 283 AA; 32718 MW; 3BA2631A5E28A993 CRC64;

Query Match 61.4%; Score 35; DB 1; Length 283;
Best Local Similarity 55.6%; Pred. No. 17;
Matches 5; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 CRRETAMAC 9
DB 267 CRPYRWVC 275

RESULT 6

```

GALE.RAT  
ID GALE.RAT STANDARD: PRT: 347 AA.  
AC P18645;  
DT 01-NOV-1990 (Rel. 16, Created)  
DT 01-NOV-1990 (Rel. 16, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE UDP-GLUCOSE 4-EPIMERASE (EC 5.1.3.2) (GALACTOMALDENASE) (UDP-  
GALACTOSE 4-EPIMERASE).  
GN GALE.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
RN (1)  
RP SEQUENCE FROM N.A.  
RX TISSUE-SKELETAL MUSCLE;  
RX MEDLINE-90384840; PubMed-2205840;  
RA Zeschniygk M., Wilcken-Bergmann B., Storzinski-Powltz A.;  
RT "cDNA from rat cells with recombinative galactose-epimerase activity  
in E. coli.";  
RL Nucleic Acids Res. 18:5289-5289(1990).  
CC -1- FUNCTION: CATALYZES TWO DISTINCT BUT ANALOGOUS REACTIONS: THE  
CC EPIMERIZATION OF UDP-GLUCOSE TO UDP-GALACTOSE AND THE  
CC EPIMERIZATION OF UDP-N-ACETYLGLUCOSAMINE TO UDP-N-  
CC ACETYLGLUCOSAMINE.  
CC -1- CATALYTIC ACTIVITY: UDP-GLUCOSE -> UDP-GALACTOSE.  
CC -1- COFACTOR: NAD.  
CC -1- PATHWAY: GALACTOSE METABOLISM.  
CC -1- SIMILARITY: BELONGS TO THE SUGAR EPIMERASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: X53949; CA37897.1; -  
DR PIR: S11323; S11223.  
DR HSP: P08147; IKVS.  
DR INTERPRO: IPR001509; -  
DR PFM: PFO1370; Epimerase; 1.  
KW isomerase; NAD; Galactose metabolism.  
FT NP\_BIND 4  
FT 35 NAD (POTENTIAL).  
SQ SEQUENCE 347 AA; 38225 MW; 625B81546E99143 CRC64;

Query Match 61.4%; Score 35; DB 1; Length 347;  
Best Local Similarity 71.4%; Pred. No. 21;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

GY 1 CRRRTAW 7  
DB 171 CRADTAW 177

RESULT 7  
PVR\_HUMAN STANDARD: PRT: 417 AA.  
ID PVR\_HUMAN  
AC P15151; P15152;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE POLIOVIRUS RECEPTOR PRECURSOR (CD155 ANTIGEN).  
GN PVS OR PVR.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
RN (1)  
RP SEQUENCE FROM N.A.  
RX MEDLINE-89168426; PubMed-2538245;  
RA Mendelsohn C.L., Wimmer E., Racanelli V.R.;  
RT "Cellular receptor for poliovirus: molecular cloning, nucleotide

RT sequence, and expression of a new member of the immunoglobulin  
RT superfamily.";  
RT Cell 56:855-865(1989).  
RN (2)  
RP REVISIONS.  
RA Racanelli V.R.;  
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.  
RN (3)  
RP SEQUENCE FROM N.A.  
RX MEDLINE-9106015; PubMed-2170108;  
RA Kolke S., Horie H., Ise I., Okitsu A., Yoshida M., Iizuka N.,  
RA Takeuchi K., Takegami T., Nomoto A.;  
RT "The poliovirus receptor protein is produced both as membrane-bound  
RT and secreted forms.";  
RL EMBO J. 9:3217-3224(1990).  
RN (4)  
RP MEDLINE-91239515; PubMed-1851992;  
RX Kolke S., Ise I., Nomoto A.;  
RT "Functional domains of the poliovirus receptor.";  
RL Proc. Natl. Acad. Sci. U.S.A. 88:4104-4108(1991).  
RN (5)  
RP METAGENESIS OF CARBOHYDRATE-LINKAGE SITES.  
RX MEDLINE-93059689; PubMed-1331527;  
RA Zibert A., Wimmer E.;  
RT "N-glycosylation of the virus binding domain is not essential for  
RT function of the human poliovirus receptor.";  
RL J. Virol. 66:7368-7373(1992).  
CC -1- FUNCTION: NOT KNOWN. USED BY POLIOVIRUS TO BIND AND ENTER THE  
CC CELL.  
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (ALPHA AND DELTA  
CC FORMS); SECRETED (BETA AND GAMMA FORMS).  
CC -1- ALTERNATIVE PRODUCTS: 4 ISOFORMS; ALPHA (SHOWN HERE), BETA, GAMMA  
CC AND DELTA; ARE PRODUCED BY ALTERNATIVE SPLICING.  
CC -1- MISCELLANEOUS: THE V-TYPE DOMAIN IS NECESSARY AND SUFFICIENT FOR  
CC VIRUS BINDING AND UPTAKE.  
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS  
CC ONE V-LIKE AND 2 C2-LIKE DOMAINS.  
CC -1- DATABASE: NAME-PROT; NOTE-CD guide CD155 entry;  
CC WWW-"<http://www.ncbi.nlm.nih.gov/prov/cd/cd155.htm>".  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: M24407; AAA36461.1; -  
DR EMBL: M24406; AAA36462.1; -  
DR EMBL: X64116; CAA45478.1; -  
DR EMBL: X64117; CAA45478.1; JOINED.  
DR EMBL: X64118; CAA45478.1; JOINED.  
DR EMBL: X64119; CAA45478.1; JOINED.  
DR EMBL: X64120; CAA45478.1; JOINED.  
DR EMBL: X64121; CAA45478.1; JOINED.  
DR EMBL: X64122; CAA45478.1; JOINED.  
DR EMBL: X64123; CAA45478.1; JOINED.  
DR PIR: S12048; RWHUPA.  
DR PIR: A43024; RWHUPA.  
DR MIM: 173850; -  
DR INTERPRO: IPR003006; -  
DR PFM: PFO0047; 19; 3.  
DR PFM: PFO0047; 19; 3.  
KW Immunoglobulin domain; Receptor; Transmembrane; glycoprotein; signal;  
FT Alternative splicing; Polymorphism.  
FT SIGNAL 1 20  
FT CHAIN 21 417  
FT DOMAIN 21 343  
FT TRANSMEM 344 367  
FT DOMAIN 368 417  
FT DOMAIN 42 130  
FT DOMAIN 159 228  
IG-LIKE C2-TYPE DOMAIN.

```

FT DOMAIN 259 319 IG-LIKE C2-TYPE DOMAIN.
FT DISULFID 49 123 BY SIMILARITY.
FT DISULFID 166 221 BY SIMILARITY.
FT DISULFID 266 312 BY SIMILARITY.
FT CARBOHYD 105 105 N-LINKED (GLCNAC. . .)
FT CARBOHYD 120 120 N-LINKED (GLCNAC. . .)
FT CARBOHYD 188 188 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 218 218 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 237 237 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 278 278 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 307 307 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 313 313 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT VARSPLIC 340 384 MISSING (IN ISOFORM BETA).
FT VARSPLIC 331 331 E -> G (IN ISOFORM GAMMA).
FT VARSPLIC 385 392 MISSING (IN ISOFORM GAMMA).
FT VARSPLIC 393 417 TEHASASA -> EHHOSCRN (IN ISOFORM DELTA).
FT VARSPLIC 67 67 MISSING (IN ISOFORM DELTA).
FT VARSPLIC 67 67 A -> T.
FT CONFLICT 340 340 /FTID-VAR_003952.
FT SEQUENCE 417 AA; 45302 MW; DISC012CEB53169B CRC64;

```

```

Query Match 61.4%; Score 35; DB 1; Length 417;
Best Local Similarity 55.6%; Pred. No. 25;
Matches 5; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

OY 1 CRETAMAC 9
DB 370 CSREVIMHC 378

```

```

RESULT 8
VNS1_AHSV9 STANDARD: PRT; 548 AA.
ID VNS1_AHSV9
AC 085967;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE NONSTRUCTURAL PROTEIN NS1.
GN 55.
OS African horse sickness virus 9 (AHSV-9) (African horse sickness virus
OC (serotype 9)).
OC Viruses; dsRNA viruses; Reoviridae; Orbivirus.
RN [1]
RP SEQUENCE FROM N.A.
RL Submitted (AUG-1993) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; 001069; AAA16200.1; -
CC INTERPRO; IPR002630; -
CC PFM; PFO1718; ORBLNS1; 1.
CC Nonstructural protein.
SO SEQUENCE 548 AA; 63377 MW; BA625012A1EF12AD CRC64;

```

```

Query Match 61.4%; Score 35; DB 1; Length 548;
Best Local Similarity 71.4%; Pred. No. 33;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
OY 3 RETAMAC 9
DB 78 ROTIMAC 84

```

```

RESULT 9
PHRL_SIGNAL STANDARD: PRT; 501 AA.
ID PHRL_SIGNAL
AC P40115;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE DEOXYRIBIDIPYRIMIDINE PHOTOLYASE (EC 4.1.99.3) (DNA PHOTOLYASE)
GN (PHOTOREACTIVATING ENZYME).
OS Sinapis alba (white mustard) (Brassica hirta).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Sinapis.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-94073224; PubMed-8252071;
RA Batschauer A.;
RT "A plant gene for photolysis: an enzyme catalyzing the repair of UV-
RT light-induced DNA damage."
RL Plant J 4:705-709(1993).
CC -1- FUNCTION: THIS ENZYME CATALYZES THE LIGHT-DEPENDENT MONOMERIZATION
CC (300-600 NM) OF CYCLOBUTYL PYRIMIDINE DIMERS (IN CIS-SYN
CC CONFORMATION), WHICH ARE FORMED BETWEEN ADJACENT BASES ON THE
CC SAME DNA STRAND, UPON EXPOSURE TO ULTRAVIOLET RADIATION.
CC -1- CATALYTIC ACTIVITY: CYCLOBUTADIPYRIMIDINE (IN DNA) - 2 PYRIMIDINE
CC RESIDUES (IN DNA).
CC -1- COFACTOR: CONTAINS 2 CHROMOPHORES: A REDUCED FLAVIN (FADH2) AND AN
CC OXIDIZED 8-HYDROXY-5-DEAZAFLAVIN (F420). BOTH CHROMOPHORES ARE
CC BOUND BY NON-COVALENT INTERACTIONS (PROBABLE).
CC -1- INDUCTION: BY VISIBLE LIGHT.
CC -1- SIMILARITY: BELONGS TO THE DNA PHOTOLYASE CLASS-1 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; X72019; CAA50898.1; -
CC INTERPRO; IPR000474; -
CC DR INTERPRO; IPR002081; -
CC DR PFM; PFO0875; DNA_photolysase; 1.
CC DR PROSITE; PS00394; DNA_PHOTOLYASES_1_1; 1.
CC DR PROSITE; PS00691; DNA_PHOTOLYASES_1_2; 1.
CC KW Lyase; Chromophore; Flavoprotein; FAD; DNA repair; DNA-binding.
SO SEQUENCE 501 AA; 57021 MW; C6B250CE0A33690DB CRC64;

```

```

Query Match 59.6%; Score 34; DB 1; Length 501;
Best Local Similarity 83.3%; Pred. No. 45;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 4 ETAMAC 9
DB 181 ETWAMC 186

```

```

RESULT 10
PYG4_SYNEL STANDARD: PRT; 251 AA.
ID PYG4_SYNEL
AC P50041;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE PHYCOBILISOME ROD-CORE LINKER POLYPEPTIDE CPCG4.
GN CPCG4.
OS Synechococcus elongatus.
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
RN [1]
RP SEQUENCE FROM N.A.

```

RA Shlitzman T., Soga M., Hirano M., Katoh S.;  
 RL Submitted (SEP-1992) to the EMBL/Genbank/DBJ databases.  
 CC -1- FUNCTION: ROD-CORE LINKER PROTEIN REQUIRED FOR ATTACHMENT OF  
 CC PHOCOCANTIN TO ALLOPHOCOCANTIN IN CORES OF PHYCOBILISOMES.  
 CC -1- FUNCTION: LINKER POLYPEPTIDES DETERMINE THE STATE OF AGGREGATION  
 CC AND THE LOCATION OF THE DISC-SHAPED PHYCOBILIPROTEIN UNITS WITHIN  
 CC THE PHYCOBILISOME AND MODULATE THEIR SPECTROSCOPIC PROPERTIES IN  
 CC ORDER TO MEDATE A DIRECTED AND OPTIMAL ENERGY TRANSFER.  
 CC -1- SUBUNIT: THE PHYCOBILISOME IS A HEMIDISCAL STRUCTURE THAT IS  
 CC COMPOSED OF TWO DISTINCT SUBSTRUCTURES: A CORE COMPLEX (THAT  
 CC CONTAINS PHYCOBILIPROTEINS) AND A NUMBER OF RODS RADIATING FROM  
 CC THE CORE.  
 CC -1- SIMILARITY: TO OTHER PHYCOBILISOME LINKER PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: D13173; BAA02463.1; -  
 CC INTERPRO: IPR001297; -  
 CC PFM: PFM00427; PBS\_Linker\_poly; 1;  
 CC PHYCOBILISOME; Photosynthesis; Multi-gene family.  
 CC KW PHYCOBILISOME; 251 AA; 29585 MW; A793142A0051CD7E CRC64;  
 CC SEQUENCE

Query Match 57.9%; Score 33; DB 1; Length 251;  
 Best Local Similarity 71.4%; Pred. No. 35;  
 Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 RETANAC 9  
 11 11 11  
 Db 45 REVIMAC 51

RESULT 11  
 VME1\_CVPS STANDARD; PRT; 262 AA.  
 ID VME1\_CVPS  
 AC P09175;  
 DT 01-MAR-1989 (Rel. 10, Created)  
 DT 01-MAR-1992 (Rel. 10, Last sequence update)  
 DT 01-AUG-1992 (Rel. 23, Last annotation update)  
 DE E1 GLYCOPROTEIN PRECURSOR (MATRIX GLYCOPROTEIN) (MEMBRANE  
 DE GLYCOPROTEIN).  
 DE M.  
 GN Porcine transmissible gastroenteritis coronavirus (strain FS772/70).  
 OS Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;  
 OC Coronaviridae; Coronavirinae.  
 RN [1]  
 RP MEDLINE-89013891; PubMed-2845226;  
 RA Britton P., Carmenes R.S., Page K.W., Garves D.J.;  
 RT "The integral membrane protein from a virulent isolate of  
 RT transmissible gastroenteritis virus: molecular characterization,  
 RT sequence and expression in *Escherichia coli*.";  
 RL Mol. Microbiol. 2:497-505(1988).  
 CC -1- FUNCTION: TARGETS THE SITE OF VIRUS MORPHOGENESIS AND MAY BE  
 CC IMPLICATED IN VIRAL PATHOGENESIS.  
 CC -1- SUBCELLULAR LOCATION: LARGELY EMBEDDED IN THE LIPID BILAYER.  
 CC -1- SIMILARITY: BELONGS TO THE CORONAVIRUS M PROTEIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: Y00560; CAA68643.1; -

DR PIR: S01914; S01914.  
 DR INTERPRO: IPR002574; -  
 DR PFM: PFM01635; Corona\_M; 1;  
 KW Matrix protein; Glycoprotein; Transmembrane; Signal.  
 FT SIGNAL 1  
 FT CHAIN 17  
 FT TRANSMEM 57 262 E1 GLYCOPROTEIN.  
 FT TRANSMEM 77 77 POTENTIAL.  
 FT TRANSMEM 83 103 POTENTIAL.  
 FT TRANSMEM 114 134 POTENTIAL.  
 FT CARBOHYD 32 32 N-LINKED (GLCNAC... ) (PROBABLE).  
 FT CARBOHYD 55 55 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CARBOHYD 251 251 N-LINKED (GLCNAC... ) (POTENTIAL).  
 SO SEQUENCE 262 AA; 29458 MW; 79E3847C6C74671F CRC64;

Query Match 57.9%; Score 33; DB 1; Length 262;  
 Best Local Similarity 66.7%; Pred. No. 37;  
 Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CRSTANAC 9  
 11 11 11  
 Db 30 CRNSTANAC 38

## RESULT 12

PRCK\_RHIME STANDARD; PRT; 536 AA.  
 ID PRCK\_RHIME  
 AC P43085;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE PHOSPHOENOLPYRUVATE CARBOXYKINASE [ATP] (EC 4.1.1.49).  
 GN PCKA.  
 OS Rhizobium meliloti (Sinorhizobium meliloti).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Rhizobiaceae; Sinorhizobium.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-S047 / 1021;  
 RX MEDLINE-95189720; PubMed-7883700;  
 RA Oesterleas M., Driscoll B.T., Finan T.M.;  
 RT "Molecular and expression analysis of the Rhizobium meliloti  
 RT phosphoenolpyruvate carboxykinase (pckA) gene.";  
 RL J. Bacteriol. 177:1452-1460(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-1021;  
 RA Galibert F., Capela D., Hubler-Barloy F., Gatius M., Batut J.,  
 RA Boistard P., Guzy J., Kahn D., Thebaud P., Goffeau A.,  
 RA Purnelle B., Pohl T., Bothe G., Schneider S., Portetelle D.,  
 RA Vandenbol M., Puhler A., Becker A., Weidner S.;  
 RL Submitted (MAR-2000) to the SWISS-PROT data bank.

CC -1- CATALYTIC ACTIVITY: ATP + OXALOACETATE -> ADP + PHOSPHOENOLPYRUVATE  
 CC + CO(2).  
 CC -1- PATHWAY: RATE-LIMITING GLUCONEOGENIC ENZYME.  
 CC -1- SIMILARITY: BELONGS TO THE PHOSPHOENOLPYRUVATE CARBOXYKINASE [ATP]  
 CC FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U15199; AAA69973.1; -  
 CC HSSP: P22259; IOEN.  
 DR INTERPRO: IPR001272; -  
 DR PFM: PFM01293; PRCK\_ATP; 1;  
 DR PROSITE: PS00532; PRCK\_ATP; 1  
 KW Gluconeogenesis; Lyase; Decarboxylase; ATP-binding.  
 FT Np\_BIND 236 243 ATP (POTENTIAL).

SO SEQUENCE 536 AA; 58119 MW; 1A0F7CE6C3EBAED CRC64;

Query Match 57.9%; Score 33; DB 1; Length 536;  
Best Local Similarity 66.7%; Pred. No. 73;  
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 CRETMAMC 9  
111111  
DB 173 CRSTVINC 181

RESULT 13

HIS2\_CANAL STANDARD; PRT; 838 AA.

AC 074712;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE HISTIDINE BIOSYNTHESIS TRIFUNCTIONAL PROTEIN (INCLUDES:  
DE PHOSPHORIBOSYL-AMP CYCLOHYDROLASE (EC 3.5.4.19); PHOSPHORIBOSYL-ATP  
DE PHOSPHOHYDROLASE (EC 3.6.1.31); HISTIDINOL DEHYDROGENASE  
DE (EC 1.1.1.23) (HDH)).  
GN HIS4.  
OS Candida albicans (Yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;  
OC anamorphic Saccharomycetales; Candida.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ATCC 64385 / 1001;  
RX MEDLINE-98451822; PubMed-9778800;  
RA Navarro-Garcia F., Perez-Diaz R., Negredo A., Pla J., Nombela C.;  
RT "Cloning and sequence of a PEK3-like gene from Candida albicans";  
RL Yeast 14:1147-1157(1998).  
CC -1- CATALYTIC ACTIVITY: 5-PHOSPHORIBOSYL-AMP + H(2)O =  
CC 5-(5-PHOSPHO-D-RIBOSYLAMINOPURIMINO)-1-(5-PHOSPHO-RIBOSYL)  
CC IMPIDAZOLE-4-CARBOXAMIDE.  
CC -1- CATALYTIC ACTIVITY: 5-PHOSPHORIBOSYL-ATP + H(2)O =  
CC 5-PHOSPHORIBOSYL-AMP + PHOSPHATE.  
CC -1- CATALYTIC ACTIVITY: L-HISTIDINOL + 2 NAD(+) = L-HISTIDINE +  
CC 2 NADH.  
CC -1- PATHWAY: SECOND, THIRD AND TENTH STEPS IN HISTIDINE BIOSYNTHETIC  
CC PATHWAY.  
CC -1- SIMILARITY: IN THE HDH DOMAIN, TO OTHER PROKARYOTIC, FUNGAL AND  
CC PLANTS HDH.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: AJ003115; CA005871.1;  
CC INTERPRO: IPR001692;  
CC INTERPRO: IPR002496;  
CC INTERPRO: IPR002497;  
CC PFAM: PR00815; Histidinol-dh; 1.  
CC PFAM: PF01502; PRA-CH; 1.  
CC PFAM: PF01503; PRA-CH; 1.  
CC PRINTS: PR00083; HOLDHGRNASE.  
CC PROSITE: PS00611; HISOL\_DEHYDROGENASE; 1.  
CC Histidine biosynthesis; Multifunctional enzyme; Hydroxylase;  
CC Oxidoreductase; NAD  
CC  
CC FT DOMAIN 1 271 PHOSPHORIBOSYL-AMP CYCLOHYDROLASE.  
CC FT DOMAIN 272 360 PHOSPHORIBOSYL-ATP PYROPHOSPHOHYDROLASE.  
CC FT DOMAIN 361 838 HISTIDINOL DEHYDROGENASE.  
CC FT ACT\_SITE 512 512 BY SIMILARITY.  
CC SEQUENCE 838 AA; 91835 MW; BFA0E97E9CA6C03 CRC64;

Query Match 57.9%; Score 33; DB 1; Length 838;  
Best Local Similarity 55.6%; Pred. No. 1,1e+02;  
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 CRETMAMC 9  
111111  
DB 272 CHRETKFTC 280

RESULT 14

XPC\_HUMAN STANDARD; PRT; 939 AA.

AC 001831;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE DNA-REPAIR PROTEIN COMPLEMENTING XP-C CELLS (XERODERMA PIGMENTOSUM  
DE GROUP C COMPLEMENTING PROTEIN) (P125).  
GN XPC OR XPC.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-54.  
RX MEDLINE-94222030; PubMed-8168482;  
RA Masutani C., Sugawara K., Yanagisawa J., Sonoyama T., Ue M.,  
RA Enomoto T., Takio K., Tanaka K., van der Spek P.J., Bootsma D.,  
RA Hoeljmakers J.H.J., Hanaoka F.;  
RT "Purification and cloning of a nucleotide excision repair complex  
RT involving the xeroderma pigmentosum group C protein and a human  
RT homologue of yeast RAD23.";  
RL EMBO J. 13:1831-1843(1994).  
RN [2]  
RP SEQUENCE OF 118-940 FROM N.A.  
RX MEDLINE-92396218; PubMed-1522891;  
RT Legeurski R.J., Peterson C.A.;  
RT "Expression cloning of a human DNA repair gene involved in Xeroderma  
RT pigmentosum group C.";  
RL Nature 359:70-73(1992).  
RN [3]  
RP REVIEW ON VARIANTS XP-C.  
RX MEDLINE-99374920; PubMed-10447254;  
RA Cleaver J.E., Thompson L.H., Richardson A.S., States J.C.;  
RT "A summary of mutations in the UV-sensitive disorders: Xeroderma  
RT pigmentosum, Cockayne syndrome, and trichothiodystrophy.";  
RL Hum. Mutat. 14:9-22(1999).  
RN [4]  
RP VARIANTS XP-C HIS-333; VAL-696 INS AND GUN-938.  
RX MEDLINE-94129623; PubMed-8298653;  
RA Li L., Bales E.S., Peterson C.A., Legeurski R.J.;  
RT "Characterization of molecular defects in Xeroderma pigmentosum group  
RT C.";  
RL Nat. Genet. 5:413-417(1993).

CC -1- FUNCTION: INVOLVED IN DNA EXCISION REPAIR. MAY PLAY A PART IN DNA  
CC DAMAGE RECOGNITION AND/OR IN ALTERING CHROMATIN STRUCTURE TO  
CC ALLOW ACCESS BY DAMAGE-PROCESSING ENZYMS.  
CC -1- SUBUNIT: HETERODIMER OF A 125 KDA SUBUNIT (P125) AND OF A  
CC 58 KDA SUBUNIT (P58).  
CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).  
CC -1- DISEASE: DEFECTS IN XPC ARE A CAUSE OF XERODERMA PIGMENTOSUM, A  
CC RARE HUMAN AUTOSOMAL RECESSIVE DISEASE CHARACTERIZED BY SOLAR  
CC SENSITIVITY, HIGH PREDISPOSITION FOR DEVELOPING CANCERS ON AREAS  
CC EXPOSED TO SUNLIGHT AND, IN SOME CASES, NEUROLOGICAL  
CC ABNORMALITIES.

CC -1- SIMILARITY: SOME, TO YEAST RAD4.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

CC -----
DR EMBL; D21089; BAA04651.1;
DR EMBL; X65024; CAA46158.1;
DR MIM; 276720;
KW DNA repair; DNA-binding; Nuclear protein; Xeroderma pigmentosum;
  Disease mutation.
FT INIT_MET 0
FT DOMAIN 29 176 GLU-RICH (ACIDIC).
FT DOMAIN 358 394 LYS-RICH (BASIC).
FT DOMAIN 389 394 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 407 430 ARG/LYS-RICH (BASIC).
FT DOMAIN 421 460 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 465 492 ARG/LYS-RICH (BASIC).
FT DOMAIN 29 33 POLY-GLU.
FT DOMAIN 123 129 POLY-GLU.
FT DOMAIN 500 506 POLY-SER.
FT VARIANT 333 333 P -> H (IN XP-C; SEVERE).
FT VARIANT 696 696 /FTID-VAR_005846.
FT VARIANT 938 938 V -> VV (IN XP-C; MILD).
FT CONFLICT 498 498 /FTID-VAR_005847.
FT CONFLICT 498 498 K -> Q (IN XP-C; MILD).
FT CONFLICT 498 498 /FTID-VAR_005848.
FT CONFLICT 498 498 V -> A (IN REF. 2).
SQ SEQUENCE 939 AA; 105849 MW; 25BD5DC0B212E2E3 CRC64;

```

```

Query Match
Best Local Similarity 57.9%; Score 33; DB 1; Length 939;
Matches 4; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

```

```

OY 1 CRRFTAW 9
DB 535 CEQEKWVC 543

```

```

RESULT 15
HA25_HUMAN
ID P01907; STANDARD; PRT; 255 AA.
AC 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-JAN-1990 (Rel. 13, Last annotation update)
DE HLA CLASS II HISTOCOMPATIBILITY ANTIGEN, DO(5) ALPHA CHAIN PRECURSOR
  (DC-1 ALPHA CHAIN).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A. (CLONE HB20).
RX MEDLINE-84168117; PubMed-6584734;
RA Aufferay C., Lillie J.W., Arnott D., Grossberger D., Kappes D.,
  Strominger J.L.;
RT "Isotypic and allelic variation of human class II
  histocompatibility antigen alpha-chain genes.";
RL Nature 308:327-333(1984).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL Outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X00452; CAA25141.1;
DR PIR; A02211; HLHUD1.
DR HSSP; P01910; IIAK.
DR INTERPRO; IPR000495;
DR INTERPRO; IPR001003;
DR INTERPRO; IPR003006;
DR PFAM; PF00993; MHC-II_alpha; 1.
DR PFAM; PF00047; Ig; 1.
DR PROSITE; PS00290; IG_MHC; 1.

```

```

KW MHC II; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 23
FT CHAIN 24 255 HLA CLASS II HISTOCOMPATIBILITY ANTIGEN,
  DO(5) ALPHA CHAIN.
FT DOMAIN 24 110 EXTRACELLULAR ALPHA-1.
FT DOMAIN 111 204 EXTRACELLULAR ALPHA-2.
FT DOMAIN 205 217 CONNECTING PEPTIDE.
FT TRANSMEM 218 240 CYTOPLASMIC TAIL.
FT DOMAIN 241 255 BY SIMILARITY.
FT DISULFID 133 189 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 104 104 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 144 144 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 255 AA; 28058 MW; 2E3F13105919571 CRC64;

```

```

Query Match
Best Local Similarity 56.1%; Score 32; DB 1; Length 255;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 2 RRETAW 7
DB 64 RRETAW 69

```

```

Search completed: February 27, 2001, 09:02:20
Job time: 116 sec

```



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2001, 09:00:00 ; Search time 20.79 Seconds  
(without alignments)  
50.739 Million cell updates/sec

Title: US-09-307-223-1

Perfect score: 57

Sequence: 1 CRRETAMAC 9

Scoring table:  
BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 374700 seqs, 117207915 residues  
Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SP archaea: \*  
2: SP bacteria: \*  
3: SP fungi: \*  
4: SP human: \*  
5: SP invertebrate: \*  
6: SP mammal: \*  
7: SP mhc: \*  
8: SP organelle: \*  
9: SP phage: \*  
10: SP plant: \*  
11: SP rodent: \*  
12: SP virus: \*  
13: SP vertebrate: \*  
14: SP unclassified: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	43	75.4	1681	4 Q9UGL1	Q9UGL1 homo sapien
2	41	71.9	119	9 Q9MC80	Q9MC80 bacterioph
3	41	71.9	263	2 Q51552	Q51552 pseudomonas
4	39	68.4	336	2 Q54021	Q54021 paracoccus
5	38	66.7	272	5 Q17954	Q17954 caenorhabd
6	38	66.7	500	2 Q9JPM5	Q9JPM5 rhodocyclu
7	38	66.7	597	2 Q9L225	Q9L225 streptomyce
8	38	66.7	804	10 Q9MAA3	Q9MAA3 arabidopsi
9	37	64.9	269	2 Q9X1H6	Q9X1H6 neisseria m
10	37	64.9	269	2 Q9JX23	Q9JX23 neisseria m
11	36	63.2	233	1 Q9YEF4	Q9YEF4 aeropyrum p
12	36	63.2	333	5 Q62354	Q62354 caenorhabd
13	36	63.2	334	4 Q9UP08	Q9UP08 homo sapien
14	36	63.2	465	5 Q18021	Q18021 caenorhabd
15	36	63.2	2055	4 Q75055	Q75055 homo sapien
16	35	61.4	237	4 Q9UMR7	Q9UMR7 homo sapien
17	35	61.4	237	4 Q9UI34	Q9UI34 homo sapien
18	35	61.4	237	4 Q9NS33	Q9NS33 homo sapien
19	35	61.4	265	2 P74612	P74612 synechocyst

## ALIGNMENTS

RESULT	ID	Q9UGL1	PRELIMINARY	PRT	1681 AA	Q9X1H6	Q9X1H6
20	35	61.4	431	4	Q9X1H6	Q9X1H6	Q9X1H6
21	35	61.4	477	10	Q65563	Q65563	Q65563
22	35	61.4	530	2	Q59327	Q59327	Q59327
23	35	61.4	653	10	Q9S9J8	Q9S9J8	Q9S9J8
24	35	61.4	661	10	Q9LX00	Q9LX00	Q9LX00
25	35	61.4	1141	12	Q82456	Q82456	Q82456
26	35	61.4	1153	5	Q93598	Q93598	Q93598
27	34.5	60.5	322	2	Q9K1Z1	Q9K1Z1	Q9K1Z1
28	34	59.6	101	4	Q9Y552	Q9Y552	Q9Y552
29	34	59.6	119	10	Q48571	Q48571	Q48571
30	34	59.6	134	9	Q9MC80	Q9MC80	Q9MC80
31	34	59.6	151	5	Q20074	Q20074	Q20074
32	34	59.6	153	2	Q82463	Q82463	Q82463
33	34	59.6	155	2	P74426	P74426	P74426
34	34	59.6	198	9	Q9XJ72	Q9XJ72	Q9XJ72
35	34	59.6	227	10	Q9LX26	Q9LX26	Q9LX26
36	34	59.6	245	4	Q9Y549	Q9Y549	Q9Y549
37	34	59.6	275	12	P88945	P88945	P88945
38	34	59.6	291	2	Q9X5V6	Q9X5V6	Q9X5V6
39	34	59.6	297	2	Q9L9E6	Q9L9E6	Q9L9E6
40	34	59.6	303	2	Q9S1K9	Q9S1K9	Q9S1K9
41	34	59.6	313	4	Q75785	Q75785	Q75785
42	34	59.6	317	4	Q15319	Q15319	Q15319
43	34	59.6	373	5	Q9XVH2	Q9XVH2	Q9XVH2
44	34	59.6	400	2	Q83342	Q83342	Q83342
45	34	59.6	421	5	Q9VH16	Q9VH16	Q9VH16

RESULT	ID	Q9UGL1	PRELIMINARY	PRT	1681 AA	Q9X1H6	Q9X1H6
Q9UGL1	Q9UGL1	Q9UGL1	Q9UGL1	Q9UGL1	Q9UGL1	Q9UGL1	Q9UGL1
AC	Q9UGL1	Q9UGL1	Q9UGL1	Q9UGL1	Q9UGL1	Q9UGL1	Q9UGL1
DT	01-MAY-2000	(Tremblrel. 13, Created)					
DT	01-MAY-2000	(Tremblrel. 13, Last sequence update)					
DT	01-JUN-2000	(Tremblrel. 14, Last annotation update)					
DE	RB-BINDING PROTEIN.						
GN	RBBP2H1A.						
OS	Homo sapiens (Human).						
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;						
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.						
OX	NCBI_TaxID=9606;						
RN	[1]						
RP	SEQUENCE FROM N.A.						
RA	Keshuba V., Protopopov A., Podowski R., Zabarovsky E.						
RT	"Isolation and chromosomal localization of a new human retinoblastoma						
RT	binding protein 2 homolog 1a (RBBP2H1A)."						
RL	Submitted (Aug-1998) to the EMBL/GenBank/DBJ databases.						
DR	EMBL; AJ243706; CAB63108.1;						
DR	INTERPRO; IPR001606;						
DR	INTERPRO; IPR001965;						
DR	PFAM; PF00628; PHD; 3.						
DR	PFAM; PF01388; ARID; 1.						
SO	SEQUENCE	1681 AA;	190117 MR;	0429FA1E7B4FB702 CRC64;			

Query Match  
Best local similarity 75.4%; Score 43; DB 4; Length 1681;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

RESULT	ID	Q9UGL1	PRELIMINARY	PRT	119 AA
Q9UGL1	Q9UGL1	Q9UGL1	Q9UGL1	Q9UGL1	Q9UGL1
AC	Q9UGL1	Q9UGL1	Q9UGL1	Q9UGL1	Q9UGL1
DT	01-OCT-2000	(Tremblrel. 15, Created)			
DT	01-OCT-2000	(Tremblrel. 15, Last sequence update)			

DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)  
 DE ORF42.  
 GN ORF42.  
 OS Bacteriophage D3.  
 OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae;  
 OC Lambda phage group.  
 OX NCBI\_TaxID=31535;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20042341; PubMed=10572124;  
 RA Gilekjan Z.A., Kropinski A.M.;  
 RT "Cloning and analysis of the capsid morphogenesis genes of Pseudomonas  
 aeruginosa bacteriophage D3; another example of protein chain mal1?";  
 RL J. Bacteriol. 181:7221-7227(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Kropinski A.M.;  
 RT "Sequence of the Temperate, Serotype-Converting, Pseudomonas  
 aeruginosa Bacteriophage D3";  
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL, AF165218; AAF80801.1;  
 SO SEQUENCE 119 AA; 13446 MW; BD61AA4D64D683AD CRC64;

Query Match  
 Best Local Similarity 71.9%; Score 41; DB 9; Length 119;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 CRRETAMAC 9  
 DB 70 CRRESDMNC 78

RESULT 3  
 ID 051552 PRELIMINARY; PRT; 263 AA.  
 AC 051552;  
 DT 01-NOV-1996 (TREMblrel. 01, Created)  
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
 DT 01-NOV-1998 (TREMblrel. 08, Last annotation update)  
 DE PPRF OROTIDINE-5'-PHOSPHAT DECARBOXYLASE.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 OX NCBI\_TaxID=287;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN-PA01;  
 RX MEDLINE=95072489; PubMed=7765522;  
 RA Strych U., Wohlfarth S., Winkler U.K.;  
 RT "Orotidine-5'-monophosphate decarboxylase from Pseudomonas aeruginosa  
 PA01: cloning, overexpression, and enzyme characterization.";  
 RL Curr. Microbiol. 39:353-359(1994).  
 DR EMBL, X65613; CAA46565.1;  
 SO SEQUENCE 263 AA; 29001 MW; D3187DB0A415CB6B CRC64;

Query Match  
 Best Local Similarity 71.9%; Score 41; DB 2; Length 263;  
 Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 CRRETAMAC 9  
 DB 240 CRRRAVAMRC 248

RESULT 4  
 ID 054021 PRELIMINARY; PRT; 336 AA.  
 AC 054021;  
 DT 01-JUN-1998 (TREMblrel. 06, Created)  
 DT 01-JUN-1998 (TREMblrel. 06, Last sequence update)  
 DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)

DE MXAC PROTEIN.  
 GN MXAC.  
 OS Paracoccus denitrificans.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;  
 OC Paracoccus.  
 OX NCBI\_TaxID=266;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN-PD1222;  
 RA Rajinders W.N.M., Harms N.;  
 RT Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL, AJ000884; CAA04379.1;  
 DR INTERPRO: IPR002035;  
 SO SEQUENCE 336 AA; 35920 MW; A84DCBDEDC3C754A CRC64;

Query Match  
 Best Local Similarity 68.4%; Score 39; DB 2; Length 336;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 RRETAMAC 9  
 DB 296 RRDILAMAC 303

RESULT 5  
 ID 017954 PRELIMINARY; PRT; 272 AA.  
 AC 017954;  
 DT 01-JAN-1998 (TREMblrel. 05, Created)  
 DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)  
 DT 01-JAN-1999 (TREMblrel. 09, Last annotation update)  
 DE LC1.2 PROTEIN.  
 GN LC1.2.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Matthews L.;  
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94150718; PubMed=7906398;  
 RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,  
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
 RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fulton L.,  
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
 RA Jones M., Kerhaw J., Kirsten J., Laister N., Latreille P.,  
 RA Lightning J., Lloyd C., McMurtry A., Laister N., Laister N.,  
 RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Shownkeen R.,  
 RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,  
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Wollman P.;  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 elegans.";  
 RL Nature 368:32-38(1994).  
 DR EMBL, Z82277; CAB05250.1;  
 SO SEQUENCE 272 AA; 30354 MW; 8BA7DB4ADC124D7A CRC64;

Query Match  
 Best Local Similarity 66.7%; Score 38; DB 5; Length 272;  
 Matches 8; Conservative 0; Mismatches 1; Indels 4; Gaps 1;

OY 1 CRRE---TAMAC 9  
 DB 190 CRRAGVATWAMC 202

RESULT 6  
 OJFPDS

```

ID 09JPD5 PRELIMINARY; PRT; 500 AA.
AC 09JPD5;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DE EXOPOLYPHOSPHATASE.
GN PFX.
OS Rhodocyclus gelatinosus (Rhodopseudomonas gelatinosa).
OC Bacteria; Proteobacteria; beta subdivision; Comamonadaceae;
OC Rubrivivax.
OX NCBI_TaxID=28068;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-IL144;
RA Nagashima K.V., Shimada K., Matsuura K.;
RT "Phylogenetic analysis of photosynthetic genes of Rhodocyclus
RT gelatinosus: Possibility of horizontal gene transfer in purple
RT bacteria". Res. 36:185-191(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-IL144;
RA MEDLINE-94132007; PubMed-830574;
RT "Primary structure and transcription of genes encoding B870 and
RT photosynthetic reaction center apoproteins from Rubrivivax
RT gelatinosus."
RL J. Biol. Chem. 269:2477-2484(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-IL144;
RA Igarashi N., Shimada K., Matsuura K., Nagashima K.V.;
RT "PHOTOSYNTHETIC GENE CLUSTER IN PURPLE BACTERIUM, RUBRIVIVAX
RT GELATINOSUS."
RL (in) Garab G. (eds.);
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-IL144;
RA Menin L., Yoshida M., Jaquinod M., Nagashima K.V., Matsuura K.,
RA Parot P., Vermeiglio A.;
RT "Dark aerobic growth conditions induce the synthesis of a high
RT midpoint potential cytochrome c8 in the photosynthetic bacterium
RT Rubrivivax gelatinosus."
RL Biochemistry 0:0-0(1999).
DR EMBL; AB034704; BAA94021.1;
SQ SEQUENCE 500 AA; 54839 MM; 4D301D723A243DEA CRC64;

Query Match 66.7%; Score 38; DB 2; Length 500;
Best Local Similarity 75.0%; Pred. No. 30;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 RRETAWAC 9
DB 360 RRELGMAC 367

```

```

RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Brown S.P., Harris D.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Redenbach M., Kleser H.M., Denapante D., Elchner A., Cullum J.,
RA Kinashl H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome."
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL138598; CAB71254.1;
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Hypothetical protein
SQ SEQUENCE 597 AA; 68079 MM; 590A9AD41C038C8 CRC64;

```

```

Query Match 66.7%; Score 38; DB 2; Length 597;
Best Local Similarity 75.0%; Pred. No. 35;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 RRETAWAC 9
DB 218 RREAARC 225

```

```

RESULT 8
OYMAA3 PRELIMINARY; PRT; 804 AA.
AC 09MAA3;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DE PUTATIVE SUGAR TRANSPORTER.
GN T12H1.13.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicaceae; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B., Wu D.,
RA Bonning C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,
RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;
RT "Arabidopsis thaliana chromosome III BAC T12H1 genomic sequence."
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC009177; AAF27022.1;
RN [2]
RP SUGAR TRANSPORT.
SQ SEQUENCE 804 AA; 88497 MM; 3D8974041CBDF627 CRC64;

```

```

Query Match 66.7%; Score 38; DB 10; Length 804;
Best Local Similarity 75.0%; Pred. No. 45;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 RRETAWAC 9
DB 460 RRETLMAC 467

```

```

RESULT 9
OYK1H6 PRELIMINARY; PRT; 269 AA.
AC 09K1H6;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)

```

DT 01-OCT-2000 (TRENBLREL. 15, Last annotation update)  
 DE HYPOTHETICAL PROTEIN NMB0175.  
 GN NMB0175.  
 OS Neisseria meningitidis (serogroup B).  
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
 OX NCBI\_TaxID=491;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MC58 / SEROGROUP B;  
 RX MEDLINE=20175755; PubMed=10710307;  
 RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,  
 RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,  
 RA Nelson W.C., Gwinn M.L., Deboy R., Peterson J.D., Hickey E.K.,  
 RA Hatt D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,  
 RA Mason T., Ciecko A., Parksey D.S., Blair E., Ciftone H., Clark E.B.,  
 RA Cotton M.D., Uitterlinden T.R., Khouri H., Qin H., Vamathevan J.,  
 RA Gill J., Scarlato V., Maignani V., Pizsa M., Grandi G., Sun L.,  
 RA Smith H.O., Fraser C.M., Moxon E.R., Rappelli R., Venter J.C.;  
 RT "Complete genome sequence of Neisseria meningitidis serogroup B strain  
 MC58.";  
 RL Science 287:1809-1815(2000).  
 DR EMBL; AE002374; AAF40632.1; -;  
 SR TIGR; NMB0175; -;  
 SO SEQUENCE 269 AA; 28425 MW; C301DBD7E19220C4 CRC64;

Query Match 64.9%; Score 37; DB 2; Length 269;  
 Best Local Similarity 62.5%; Pred. No. 26;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 RETAMAC 9  
 DB 183 RKTVMAC 190

RESULT 10  
 ID 09JX23 PRELIMINARY; PRT; 269 AA.  
 AC 09JX23;  
 DT 01-OCT-2000 (TRENBLREL. 15, Created)  
 DT 01-OCT-2000 (TRENBLREL. 15, Last sequence update)  
 DE PUTATIVE INTEGRAL MEMBRANE PROTEIN.  
 GN NMA0093  
 OS Neisseria meningitidis (serogroup A).  
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
 OX NCBI\_TaxID=65699;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-22491 / SEROGROUP A / SEROTYPE 4A;  
 RX MEDLINE=20222556; PubMed=10761919;  
 RA Parthill J., Achman M., James K.D., Bentley S.D., Churcher C.,  
 RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,  
 RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holtroyd S.,  
 RA Jags R., Leather S., Moule S., Mungall K., Quail M.A.,  
 RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,  
 RA Whitehead S., Spratt B.G., Barrall B.G.;  
 RT "Complete DNA sequence of a serogroup A strain of Neisseria  
 meningitidis 22491.";  
 RL Nature 404:502-506(2000).  
 DR EMBL; AL162753; CAB83409.1; -;  
 SO SEQUENCE 269 AA; 28455 MW; C301DBD7E19220C4 CRC64;

Query Match 64.9%; Score 37; DB 2; Length 269;  
 Best Local Similarity 62.5%; Pred. No. 26;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 RETAMAC 9  
 DB 183 RKTVMAC 190

RESULT 11  
 ID 09YEF4 PRELIMINARY; PRT; 233 AA.  
 AC 09YEF4;  
 DT 01-NOV-1999 (TRENBLREL. 12, Created)  
 DT 01-NOV-1999 (TRENBLREL. 12, Last sequence update)  
 DE 01-MAY-2000 (TRENBLREL. 13, Last annotation update)  
 GN HYPOTHETICAL 24.3 KDA PROTEIN APE0622.  
 OS Aeropyrum pernix.  
 OC Archaea; Crenarchaeota; Desulfurococcales; Desulfurococcaceae;  
 OX Aeropyrum.  
 OX NCBI\_TaxID=56666;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K1;  
 RX MEDLINE=99310339; PubMed=10382966;  
 RA Karabayasi Y., Hino Y., Horikawa H., Yamazaki S., Halkawa Y.,  
 RA Jinno K., Takahashi M., Sekine M., Baba S., Anai A., Kosugi H.,  
 RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,  
 RA Yamazaki J., Kusuda S., Funahashi T., Tanaka T., Kudoh Y.,  
 RA Nomura N., Sako Y., Kikuchi H.;  
 RT "Complete genome sequence of an aerobic hyper-thermophilic  
 crenarchaeon, Aeropyrum pernix K1.";  
 RL DNA Res. 6:83-101(1999).  
 DR EMBL; AP00060; BAA9592.1; -;  
 KW Hypothetical protein.  
 SO SEQUENCE 233 AA; 24287 MW; FDADED3BED3DAEB4 CRC64;

Query Match 63.2%; Score 36; DB 1; Length 233;  
 Best Local Similarity 55.6%; Pred. No. 35;  
 Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 CRRETAMAC 9  
 DB 164 COAPAMAC 172

RESULT 12  
 ID 062354 PRELIMINARY; PRT; 323 AA.  
 AC 062354;  
 DT 01-AUG-1998 (TRENBLREL. 07, Created)  
 DT 01-AUG-1998 (TRENBLREL. 07, Last sequence update)  
 DE 01-MAY-2000 (TRENBLREL. 13, Last annotation update)  
 GN T02G6.4 PROTEIN.  
 GN T02G6.4.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Pelodermidae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA McClay K.;  
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94150718; PubMed=7906398;  
 RA Wilson R., Alnsough R., Anderson K., Baynes C., Betks M.,  
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
 RA Jones M., Keshaw J., Kirsten J., Lister N., Latreille P.,  
 RA Lightning J., Lloyd C., Murray A., Mortimore B., O'Callaghan M.,  
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shonkhen R.,  
 RA Smalton N., Smith A., Sonhammer E., Straden R., Sulston J.,  
 RA Thierly-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
 RA Watson A., Weinstock L., Wilkinson-Spratt J., Wohlman P.,  
 RT "2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 elegans.";  
 RL Nature 368:32-38(1994).

DR EMBL: 281583; CAB04669.1;  
 DR INTERPRO: IPR001064;  
 DR PROSITE: PS00225; CRYSTALLIN\_BETAGAMMA: UNKNOWN.1.  
 SQ SEQUENCE 323 AA; 37543 MW; 738D17745079C6D CRC64;

Query Match  
 Best Local Similarity 63.2%; Score 36; DB 5; Length 323;  
 Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 1 CRRCTAMAC 9  
 DB 272 CRRCSMWC 280

RESULT 13  
 OYUFO8 PRELIMINARY; PRT: 334 AA.

DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
 DE HYPOTHETICAL 37.3 KDA PROTEIN (FRAGMENT).  
 GN DKFPA34E0335.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=TESTIS;  
 RA Poustka A., Klei M., Neues H.W., Gaassenhuber J., Wiemann S.;  
 RL EMBL: AL117402; CAB55903.1;  
 DR SUBLMIT (SEP-1999) to the EMBL/GenBank/DBJ databases.  
 KW Hypothetical protein.  
 FT NON-TER  
 SQ SEQUENCE 334 AA; 37314 MW; DB055F79DB0F3FD CRC64;

Query Match  
 Best Local Similarity 63.2%; Score 36; DB 4; Length 334;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 CRRCTAM 7  
 DB 175 CRRDTLM 181

RESULT 14  
 ID 018021 PRELIMINARY; PRT: 465 AA.

DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-JAN-1999 (TREMBLrel. 09, Last annotation update)  
 DE T04A11.10 PROTEIN.  
 GN T04A11.10.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Peloderidae; Caenorhabditis.  
 OX NCBI\_TaxID=6219;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Rottow K.;  
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RE SEQUENCE FROM N.A.  
 RX MEDLINE: 94150713; PubMed: 7906198;  
 RA Wilson R., Ah-Cowth R., Anderson K., Haynes C., Berts M.,  
 Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
 Crexson A., Dea S., Du Z., Durbin R., Favello A., Fulton L.,  
 Gardner A., Gibson P., Hawking T., Hillier L., Jier M., Johnston L.,  
 Jones M., Kershaw J., Kirsten J., Lalster N., Latreille P.,

RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
 RA Parsons J., Percy C., Rifken L., Roopa A., Saunders D., Showkseen R.,  
 RA Smailson N., Smith A., Sonhammer E., Staden R., Sulston J.,  
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.,  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 elegans."  
 RL Nature 368:32-38(1994).  
 DR EMBL: 283123; CAB05613.1;  
 SQ SEQUENCE 465 AA; 54036 MW; E4FA7926A5AB35C CRC64;

Query Match  
 Best Local Similarity 63.2%; Score 36; DB 5; Length 465;  
 Matches 4; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 CRRCTAMAC 9  
 DB 454 CRRCSMWC 462

RESULT 15  
 ID 075055 PRELIMINARY; PRT: 2055 AA.

DT 01-NOV-1998 (TREMBLrel. 08, Created)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
 DE KIA0467 PROTEIN (FRAGMENT).  
 GN KIA0467.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=BRIN;  
 RX MEDLINE=98116662; PubMed=9455484;  
 RA Seki N., Ohira M., Nagase T., Ishikawa K., Miyajima N., Nakajima D.,  
 RA Nomura N., Ohara O.;  
 RT "Characterization of cDNA clones in size-fractionated cDNA libraries  
 from human brain."  
 RL DNA Res. 4:345-349(1997).  
 DR EMBL: AB007936; BAA32312.1;  
 DR INTERPRO: IPR001424;  
 DR PROSITE: PS00087; SOD\_CU\_ZN\_1; UNKNOWN.1.  
 FT NON-TER  
 SQ SEQUENCE 2055 AA; 226004 MW; F3C0C252BFB4E5A CRC64;

Query Match  
 Best Local Similarity 63.2%; Score 36; DB 4; Length 2055;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 CRRCTAM 7  
 DB 1855 CRRDTLM 1861

Search completed: February 27, 2001, 09:01:21  
 Job time: 81 sec

